

Conference	The 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025), November 20–22, 2025, National Taiwan University, Taipei, Taiwan
Organizer	Association for Computational Linguistics and Chinese Language Processing (ACLCLP), Taiwan's national academic society; ROCLING held annually since 1988
Paper	A Preliminary Study of RAG for Taiwanese Historical Archives
Authors	Claire Lin*, Bo-Han Feng*, **Xuanjun Chen***, Te-Lun Yang, Hung-yi Lee, Jyh-Shing Roger Jang (*equal contribution; co-first author)
Selection	Only archival papers were eligible: 1 of the 36 archival papers accepted to the main conference (1 / 36 = 2.8%) · Accepted Papers ↗



ROCLING 2025
BEST PAPER AWARD

Presented to

Claire Lin, Bo-Han Feng, Xuanjun Chen,

Te-Lun Yang, Hung-yi Lee, and Jyh-Shing Roger Jang

For the paper entitled

A Preliminary Study of RAG for Taiwanese Historical Archives

ROCLING 2025 General Chairs
Yun-Nung Chen, Hung-Yi Lee and Pu-Jen Cheng, National Taiwan Univisity
The 37th Conference on Computational Linguistics and Speech Processing at Taipei, Taiwan on November 20-22, 2025

(a) Award certificate (scan, rotated upright); name in the orange box

A Preliminary Study of RAG for Taiwanese Historical Archives

Claire Lin, Bo-Han Feng, Xuanjun Chen, Te-Lun Yang, Hung-Yi Lee, Jyh-Shing Roger Jang

Abstract

Retrieval-Augmented Generation (RAG) has emerged as a promising approach for knowledge-intensive tasks. However, few studies have examined RAG for Taiwanese Historical Archives. In this paper, we present an initial study of a RAG pipeline applied to two historical Traditional Chinese datasets, Fort Zeelandia and the Taiwan Provincial Council Gazette, along with their corresponding open-ended query sets. We systematically investigate the effects of query characteristics and metadata integration strategies on retrieval quality, answer generation, and the performance of the overall system. The results show that early-stage metadata integration enhances both retrieval and answer accuracy while also revealing persistent challenges for RAG systems, including hallucinations during generation and difficulties in handling temporal or multi-hop historical queries.

PDF | Cite | Search | Fix data

Anthology ID: 2025.rocling-main.6
Volume: Proceedings of the 37th Conference on Computational Linguistics and Speech Processing (ROCLING 2025)
Month: November
Year: 2025
Address: National Taiwan University, Taipei City, Taiwan

(b) Proceedings entry · ACL Anthology ↗